

# STRUCTURAL INSIGHTS INTO THE INHIBITION OF *BACILLUS ANTHRACIS* SPORULATION BY A NOVEL CLASS OF NON-HEME GLOBIN SENSOR DOMAINS

Gudrun R. Stranzl<sup>1</sup>, Eugenio Santelli<sup>1</sup>, Laurie A. Bankston<sup>1</sup>, Chandra La Clair<sup>2</sup>, Andrey Bobkov<sup>1</sup>, Robert Schwarzenbacher<sup>1,4</sup>, Adam Godzik<sup>1</sup>, Marta Perego<sup>2</sup>, Marcin Grynberg<sup>1,3</sup> and Robert C. Liddington<sup>1</sup>

From the Infectious & Inflammatory Disease Center, Sanford-Burnham Medical Research Institute<sup>1</sup>, La Jolla, CA 92037, USA; The Scripps Research Institute, Department of Molecular and Experimental Medicine<sup>2</sup>, La Jolla, CA 92037, USA; and the Institute of Biochemistry and Biophysics PAS<sup>3</sup>, Pawinskiego 5A, 02-106 Warsaw, Poland

Running Title: Structure of a non-heme globin sensor domain

Address Correspondence to: Robert C. Liddington, Sanford-Burnham Medical Research Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA. Tel: 858-646-3136 Fax: 858-795-5225 Email: [rliddington@sanfordburnham.org](mailto:rliddington@sanfordburnham.org)

Pathogenesis by *Bacillus anthracis* requires coordination between two distinct activities: plasmid-encoded virulence factor expression (which protects vegetative cells from immune surveillance during outgrowth and replication), and chromosomally-encoded sporulation (required only during the final stages of infection). Sporulation is regulated by at least 5 sensor histidine kinases that are activated in response to various environmental cues. One of these kinases, BA2291, harbors a sensor domain that has ~35% sequence identity with 2 plasmid proteins, pXO1-118 and pXO2-61. Since over-expression of pXO2-61 (or pXO1-118) inhibits sporulation of *B. anthracis* in a BA2291-dependent manner, and pXO2-61 expression is strongly upregulated by the major virulence gene regulator, AtxA, it was suggested that their function is to titrate out an environmental signal that would otherwise promote untimely sporulation. To explore this hypothesis, we determined crystal structures of both plasmid-encoded proteins. We found that they adopt a dimeric globin fold, but, most unusually, do not bind heme. Instead, they house a hydrophobic tunnel and hydrophilic chamber that are occupied by fatty acid, which engages a conserved arginine and chloride ion via its carboxyl head-group. In vivo, these domains may therefore recognize changes in fatty acid synthesis, chloride ion concentration, and/or pH. Structure-based comparisons with BA2291 suggests that it binds ligand and

dimerizes in an analogous fashion, consistent with the titration hypothesis. Analysis of newly sequenced bacterial genomes points to the existence of a much broader family of non-heme, globin-based sensor domains, with related but distinct functionalities, that may have evolved from an ancestral heme-linked globin.

Fully virulent *B. anthracis* carries two large plasmids, pXO1 and pXO2, that are responsible for the production of its major virulence factors, anthrax toxin and the poly-γ-D-glutamic acid capsule (1-3). Following host-triggered germination of *B. anthracis* spores, toxin and capsule enable the bacilli in their vegetative form to evade the host's immune system and replicate rapidly in the lymphatic system and bloodstream. If the infection is not treated at an early stage, toxemia and septicemia leading to host death may rapidly follow (4-7).

Sporulation is required for long-term survival of *B. anthracis* following host death, but the process must be carefully coordinated with toxin expression and the progress of the infection, since sporulating cells (in contrast to encapsulated vegetative cells) are susceptible to host defenses. Given that sporulation is directed primarily by chromosomal genes, close coordination with the regulatory elements encoded by plasmid genes is required for effective pathogenesis.

Indeed, there is much evidence for such "cross-talk", although the overall picture is far

from clear (reviewed in (8)). For example, the pXO1-encoded protein, AtxA, regulates synthesis of toxin (also pXO1-encoded) as well as capsule (pXO2-encoded) (9-12). AtxA is also part of a regulatory network for S-layer synthesis (a further defensive layer distinct from the capsule and peptidoglycan cell wall), a task performed by chromosomal proteins (13). In turn the synthesis of AtxA is regulated by the chromosomally-encoded transcription factor Spo0A, the master regulator of sporulation (14,15), which completes a regulatory link between sporulation and virulence factor expression.

Spo0A is activated by phosphorylation to induce or repress transcription of genes required or not required for sporulation, respectively (16). The regulatory pathway controlling Spo0A is more complex than most two-component signal transduction systems. In *B. anthracis* this pathway includes at least 5 (chromosomally-encoded) sensor histidine kinases that are capable of inducing sporulation (17), as well as several aspartyl phosphatases, one of them encoded by the pXO1 virulence plasmid, that inhibit sporulation (18,19).

Sensor histidine kinases are the primary sensors of environmental cues, and form a large family of signaling proteins in both Gram-positive and Gram-negative bacteria. They have a modular architecture comprising at least a “sensor” domain and a catalytic domain (which includes the phosphorylatable domain DHp and the ATP-binding domain) that autophosphorylates on a histidine residue in response to sensor domain activation. Phosphohistidine is a high-energy species that transfers its phosphoryl group to an aspartic acid residue on the downstream effector (for reviews see (20)).

BA2291 is one of the most active kinases in promoting sporulation in *B. anthracis* (17). It also appears to be unique in using GTP rather than ATP as its energy source for phosphorylation (21), and orthologs are found in most members of the *B. cereus* group (a subfamily of the genus *Bacillae*, which includes *B. anthracis* and *B. thuringiensis* (an insect pathogen), but not *B. subtilis*). Two plasmid-encoded proteins, pXO1-118 and pXO2-61, express “sensor-only” domains that share ~35% sequence identity with BA2291 (22) and are only found in *B. anthracis* and certain strains of *B. cereus* that harbor similar plasmids. The *pXO1-118* gene lies next to and is divergently transcribed from the *atxA* gene, while *pXO2-61* lies within the region directing capsule synthesis. A microarray study found that transcription of *pXO2-61* was strongly upregulated when the *atxA* gene was

present (12); and overexpression of pXO2-61 was found to reduce sporulation of *B. anthracis* in a BA2291-dependent manner (22). Expression of BA2291 in a *B. subtilis* model induced sporulation when expressed at lower copy levels, and this was repressed by co-expression of either pXO1-118 or pXO2-61. However, higher levels of BA2291 expression led to repression of sporulation, suggesting that when the activating signal is in limited supply, the kinase activity is reversed, and BA2291 acts as a phosphatase. These observations led to a model in which the plasmid-encoded sensor domains modulate BA2291 activity by titrating the sporulation signal, thereby preventing premature sporulation. This model completes a regulatory link between virulence factor expression and sporulation (22).

In order to explore the molecular mechanisms of BA2291 regulation, and the nature of the sporulation signal, we determined the crystal structures of pXO1-118 and pXO2-61 at high resolution. We show that they adopt a dimeric globin fold, but, most unusually, do not bind heme. We demonstrate that they bind fatty acid and a halide ion (most likely chloride) in a central cavity, and show that the key structural features of ligand recognition and dimerization are conserved in a large family of kinases found in *Bacilli* and related species. This suggests that they recognize the same environmental cue(s), and support the titration mechanism for the sensor-only domains. Recognizable homologs are also found in a number of distinct bacterial phyla, and our analysis points to an evolution of the globin family into a “heme-free” group of versatile environmental sensors.

## EXPERIMENTAL PROCEDURES

*Cloning, expression and purification-* The plasmid for *E. coli* over-expression of *B. anthracis* ORF118 was obtained by cloning the PCR-amplified coding sequence using oligonucleotides BaORF1185’Nde (5’-GAGTGGACATATGGAAGCAACAAAACG-3’) and BaORF1183’Bam (5’-CTATAGGATCCAAAAATTTCAAGGTG-3’) into plasmid pET28a (Stratagene) digested with *Nde*I and *Bam*HI. The coding sequence for amino acids 1 to 146 of BA2291 was amplified using oligonucleotides BaKin5’Nde (5’-TATTCGTCATATGGAAATGGAGGGAATG-3’) and BaKinXho2 (5’-TTTCCCTCGAGTTTATAATATAATTTCCGAGTAT-3’) and the fragment was cloned in

pEt28a digested with *NdeI* and *XhoI*. In the case of pXO2-61, a synthetic gene was purchased from GenScript Co., NJ, USA and subcloned in pET28 as described for pXO1-118. Expression was obtained in *E. coli* BL21(DE3) grown in LB medium after induction with 0.1 mM IPTG for 4 h at 32°C. All proteins included a cleavable 6-His tag, and were purified from the soluble fraction by Nickel affinity chromatography on a His trap chelating column (Pharmacia), followed by thrombin removal of the tag, and size exclusion on Superdex75 or Superdex200 columns (Amersham, Pharmacia). The apparent MW in each case was consistent with a dimer. pXO1-118 was stored frozen in 20 mM Tris HCl buffer pH 7.4, 1 M NaCl, 50  $\mu$ M KCl, 5 mM dithiothreitol. pXO2-61 and BA2291 sensor domain required the addition of 500 mM NaCl to keep the proteins stable for freeze-thawing and long-term storage. For pXO1-118, selenomethionine-labeled protein was purified using a similar protocol, except that cells were grown in minimal media supplemented with SeMet (23). The MW of all proteins was confirmed by SDS-PAGE and MALDI-TOF mass spectrometry.

*Crystallization, data collection, and structure solution*- Native and selenomethionine-labeled pXO1-118 were crystallized by sitting or hanging drop vapor diffusion at room temperature by mixing 3  $\mu$ l of precipitant solution (40% (v/v) PEG-300, 100 mM Tris-HCl pH 5.4, 5% (w/v) PEG-1000) and 3  $\mu$ l of protein solution at 14 mg/ml. Rod-shaped crystals grew within 3 days. They belong to space group P3<sub>2</sub>21 with cell dimensions a=89.9 Å, c=35.3 Å. One native and one SeMet data set at the Se absorption edge were collected at SSRL beamline 9-2 and NSLS beamline X26C, respectively, at 100 K. Diffraction images were processed and scaled with the HKL package (24). SOLVE (25) located four Se sites, leading to a set of initial phases with Figure of Merit (FOM) of 0.32. Density modification increased the FOM to 0.60, and automatic model-building using RESOLVE (26) resulted in a 77% complete model. Further model-building was carried out in O (27), and the structure refined with REFMAC5 (28) and simulated-annealing using CNS (29). The final model for pXO1-118 contains a single domain (residues 1-150), three non-native N-terminal residues, one molecule of undecanoic acid, 95 water molecules and one Cl<sup>-</sup> ion. The final R<sub>WORK</sub>=0.181 and R<sub>FREE</sub>=0.225 for data from 80-1.76 Å resolution. The dimer is generated by rotation of the monomer about a crystallographic dyad.

pXO2-61 was crystallized by the microbatch method under paraffin oil. Crystals were obtained in 2 days using a buffer containing 1 M NaI, 20% (v/v) PEG 3350 100mM Tris-HCl pH 7.5. Crystals belong to space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> with unit cell a=44.1, b=62.6 and c= 124.7 Å. Data were collected at the SSRL to 1.49 Å resolution, and processed with the HKL package (24). The structure was solved by molecular replacement using the refined pXO1-118 structure as the search model. Model building and refinement were carried out in O (27) and REFMAC5 (28). The final model has R<sub>WORK</sub>= 17.7 % and R<sub>FREE</sub>=20.9 % for data from 60-1.49 Å resolution. The asymmetric unit contains two molecules forming a dimer (residues 5-136 for each molecule), 364 water molecules, 26 I<sup>-</sup> ions and 8 Na<sup>+</sup> ions. The solvent content is 56.7%. Data collection and refinement statistics are summarized in Table 1. The stereochemical quality of both models was confirmed using PROCHECK (30). The coordinates and structure factors for pXO2-61 and pXO1-118 have been deposited with the PDB (codes 3pmc and 3pmd).

*Gas Chromatography-Mass Spectrometry (GC-MS)*- 200  $\mu$ l of chloroform was added to 0.1-1 ml 10 mg/ml sensor domain in solution. The resulting two-phase system was sonicated for 10 min, incubated at 70°C for 1 h, and then centrifuged. The organic phase was separated by syringe. For pXO2-61, the carboxylic acid group of the fatty acid was verified by using 20  $\mu$ l bis-trimethylsilyltrifluoroacetamide and 20  $\mu$ l pyridine incubated for 1.5 h at 65°C. Samples were evaporated to dryness under a stream of N<sub>2</sub> and reconstituted in 100  $\mu$ l methylene chloride, prior to analysis with GC-MS (Scripps Center for Mass Spectrometry, CA, USA).

*Isothermal Titration Calorimetry (ITC)*- ITC was performed using a VP-ITC calorimeter from Microcal (Northampton, MA). 8  $\mu$ l of fatty acid solution (1.6-2.6 mM) was injected into the cell containing 100  $\mu$ M protein (pXO1-118 or pXO2-61) in 20 mM Tris pH 7.4 and either 500 mM or 1 M NaCl, respectively. All titrations were performed at 23°C, and each experiment involved 37 injections. Myristic acid (n-C14:0) and palmitic acid (n-C16:0) were purchased from Sigma-Aldrich, MO, USA. 12-methyltetradecanoic acid (anteiso-C15:0) and 13-methyltetradecanoic acid (iso-C15:0) were purchased from Indofine Chemical Co, NJ, USA. Palmitoleic acid was purchased from Fluka. Data were analyzed using Microcal Origin software provided by the manufacturer.

**Analytical ultracentrifugation-** Sedimentation equilibrium experiments were performed in a ProteomeLab XL-I (BeckmanCoulter) AUC. Protein samples (pXO2-61 or BA2291 sensor domain) in 20 mM Pipes pH 7.5 and 500 mM NaCl were loaded at concentrations of 0.5, 0.167 and 0.056 mg/ml in 6-channel equilibrium cells, and spun in an An-50 Ti 8-place rotor at 25,000 rpm for 24 h at 20°C. Data were analyzed with HeteroAnalysis software (J.L. Cole and J.W. Lary, University of Connecticut). In each case, an ideal equilibrium model gave a convincing fit for a monomer-dimer equilibrium in solution, with dimer MW values of 32.5 kDa (pXO2-61) and 33.8 kDa (BA2291 sensor domain), no evidence for higher oligomers, and dimerization Kds of 0.65  $\mu$ M and 0.83  $\mu$ M, respectively.

**Yeast Two-hybrid analysis-** The yeast two-hybrid system (Clontech) was used to explore interactions between pXO1-118 and AtxA. Coding genes were singly cloned into the bait plasmid (pGBT9) and prey plasmid (pGAD424). Assays were performed in the yeast strain AH109. We detected interaction when pXO1-118 was present on both pGBT9 and pGAD424, consistent with homodimer formation in yeast cells. There was no evidence for AtxA-pXO1-118 interactions (data not shown).

**Bioinformatics analysis-** Homologs of sensor domains were found using CS-BLAST ([http://toolkit.tuebingen.mpg.de/cs\\_blast](http://toolkit.tuebingen.mpg.de/cs_blast)).

## RESULTS

**Crystal and solution structures of pXO1-118 and pXO2-61.** We solved the crystal structures of pXO1-118 and pXO2-61 at 1.76 Å and 1.49 Å resolution, respectively (See Methods, Table I, Figs. 1 and 2). The asymmetric unit of pXO1-118 contains a single sensor domain which adopts the globin fold. Although it does not bind heme, we follow the standard nomenclature for hemoglobins: helices A, B, E, F, G, H are present in pXO1-118, while helices C and D are replaced by an ordered loop ("BE"). A dimer is formed across a crystallographic dyad, mediated by the packing of the G and H helices from apposing monomers, forming an antiparallel, left-handed four-helix bundle (Fig. 1) that buries a large interface ( $\sim 3500$  Å<sup>2</sup>).

As expected, the structure of pXO2-61 (Fig 2) is very similar to that of pXO1-118 in both tertiary fold and dimer organization. The asymmetric unit in this case contains a dimer, and the monomers superpose with an RMS deviation of 0.43 Å for

C $\alpha$  carbons. The most obvious external difference arises from a shorter C-terminal helix, leading to a more compact shape and a decrease in the dimerization interface, to 2200 Å<sup>2</sup>. The dimers of pXO1-118 and pXO2-61 superpose with an RMS difference of 1.2 Å for 264 C $\alpha$  carbons (0.85 Å for pairwise comparisons of single domains).

By analytical ultracentrifugation (AUC) we found that pXO2-61 forms dimers in solution at  $\geq \mu$ M concentrations (dimer Kd = .6  $\mu$ M). We also purified the sensor domain from BA2291, and found that it also forms dimers in solution, with a similar Kd ( $\sim 0.8$   $\mu$ M). (Suppl. Fig 1). In addition, we demonstrated that pXO1-118 forms dimers *in vivo*, as evidenced in a yeast 2-hybrid system (see Experimental Procedures).

The dimer interfaces of pXO1-118 and pXO2-61 include upper and lower hydrophobic regions that are more closely packed; while in the central section the helices diverge, and the interface is chiefly hydrophilic, comprising a large number of direct and water-mediated hydrophilic interactions (Fig. 1C). We found a total of 34 (pXO1-118) and 20 (pXO2-61) solvent molecules buried at distinct locations in this space. pXO1-118 has an additional interface formed by the ends of the longer C-terminal helices, which form a short antiparallel cross-over  $\beta$ -sheet with clusters of Tyr and Lys residues. Given the high homology between the BA2291 sensor domain and the plasmid domains (35% identity, 59% similarity) and the existence of a dimer in all cases, it is reasonable to assume that BA2291 will share a very similar tertiary and quaternary organization.

**pXO1-118 and pXO2-61 are non-heme globins.** 3-dimensional comparisons of the plasmid-encoded sensor domains using the DALI server (31) identified many structures with the globin fold that superpose with RMS differences in the 2.3-2.7 Å range for  $\sim 150$  C $\alpha$ s (Suppl. Table I). However, the PDB database contains no close sequence homologs (identities range from 8 to 14%). Moreover, most proteins with the globin fold have an Fe-heme cofactor sandwiched between the E and F helices. The most similar structures (in both tertiary and quaternary organization) are two bacterial (*B. subtilis* and *Geobacter*) heme-containing oxygen sensor domains, which form dimers via a similar pairing of the G and H helices. There is one globin structure that lacks co-factor altogether: the *B. subtilis* stress response regulator RsbR (32); in this case, the dimerization helices (G and H) bend inwards, eliminating the co-factor cavity. Mammalian hemoglobins have closely related

tertiary folds, but only one, the recently discovered cytoglobin, has a related dimeric organization (33). The packing in the dimeric bacterial hemoglobin from *Vitreoscilla* sp. (PDB 2VHB) (34) is more closely related to the mammalian hemoglobins.

The cyanobacter phycocyanins also adopt a globin-like fold (35); they are electron transfer proteins involved in photosynthesis, and bind porphyrin-like moieties via cysteine-mediated thioether bonds. However, they have large N-terminal extensions, and their quaternary organization is quite distinct from any of the hemoglobins.

*pXO1-118 and pXO2-61 contain a central cavity.* An overlay of pXO1-118 and the *B. subtilis* oxygen-sensor domain (36) illustrates the overall similarity in secondary and tertiary folds (Fig. 3A) and shows how helix E of pXO1-118 rotates and bends in order to pack more closely against helix F, partially filling the space that is occupied by heme in the oxygen sensor. The short rigid BE loop (which replaces the CD helix/turn) packs against the FG turn, stabilizing this conformation (see below). Oxygen-binding hemoglobins are linked to the heme via a “proximal histidine” at the 8<sup>th</sup> position of helix F (F8). In pXO1-118 (and pXO2-61), there is no similarly located histidine, and, as expected, attempts to reconstitute the proteins with hemin were unsuccessful (data not shown). Remarkably, DALI-based structural alignment places pXO1-118 residue Arg74 at the F8-heme location (Fig. 3B). A cross-section through the central cavity (Fig. 3C) further demonstrates the steric mismatch between heme and the pXO1-118 cavity. As discussed below, Arg74 is an invariant residue that appears to play a key role in binding an alternative cofactor (fatty acid - see Fig. 3B), and thus may be considered a structural and functional analog of the proximal histidine. Note that in RsbR, DALI-based structural alignment shows that helix F is truncated such that there is no analog of F8 (Fig 3D), consistent with its being cofactor-free.

*Crystals of pXO1-118 contain a buried fatty acid.* The central cavity runs roughly parallel with helices E and G, with a contour length of ~20 Å (Fig. 4). It comprises a narrow (~6 Å diameter) tunnel lined with hydrophobic and aromatic residues, capped at one end by residue Phe19, which appears to act as a gate, adopting conformations that either open or close the hydrophobic entrance to the tunnel. The other end of the tunnel opens into a hydrophilic chamber, which is sealed from bulk solvent by a “canopy” created by the packing of the BE and FG turns. It

includes water molecules as well as a heavier anion, which we believe to be chloride (see below).

In pXO1-118, the cavity is also occupied by continuous worm-like electron density that ends in a symmetric bifurcation, consistent with the presence of a fatty acid (Fig. 4a). Using mass spectrometry (GC-MS), we determined that the major chloroform-soluble non-protein component in the crystals is palmitic (hexadecanoic) acid, presumably derived from the cell wall of *E. coli* (37) during expression or cell lysis (Suppl. Fig. 2). The tunnel is long enough to completely bury tetra-decanoic acid (with the Phe19 gate in the open position), and additional methylene groups would presumably extrude into solvent. We searched for an authentic *Bacillus* cofactor by incubating *E.coli*-purified pXO1-118 with crude *B. cereus* cell extract. The protein was then repurified and its crystal structure determined. However, we observed no significant difference in the cofactor density (data not shown), suggesting that a similar molecule binds *in vivo*.

GC-MS analysis demonstrated the presence of fatty acid in *E.coli*-purified pXO2-61 as well as the BA2291 sensor domain (Suppl. Fig. 2). However, it is not present in the crystals of pXO2-61 (BA2291 failed to crystallize), most likely due to the different crystallization conditions. Thus, pXO1-118 crystallized in sodium chloride, while pXO2-61 required the presence of 1.0 M sodium iodide. Iodide is a strongly chaotropic agent that disfavors complex formation and increases the solubility of hydrophobic moieties (38), consistent with the expulsion of fatty acid from the protein. The higher pH required for crystal growth of pXO2-61 (pH 7.5 vs. 5.4) may also disfavor binding (see below). Notwithstanding, pXO2-61 binds fatty acids in solution (in the absence of iodide) with an affinity similar to that of pXO1-118 (see below). Despite the absence of fatty acid, a nearly identical tunnel and canopy are observed in crystals of pXO2-61. Weak density observed within the tunnel has been modeled as solvent/buffer; it is quite distinct from that in pXO1-118, and the density stops well short of the Phe19 gatekeeper, which adopts an ordered, closed conformation.

*A conserved hydrophilic chamber engages the fatty acid head-group and a halide ion.* The BE and FG turns pack closely together at the top of the domain, forming an extended salt-bridge/H-bonded network (“canopy”) that seals the chamber from bulk solvent, and engages the fatty acid carboxyl head-group and a tightly bound halide ion (Fig. 4). A prominent feature of the canopy is a motif, <sup>69</sup>KIAXER<sup>74</sup>, at the end of the F helix,



which is invariant among the sensor domain orthologs within the *B. cereus* group. Thus, Lys69 and Glu73 form part of a salt-bridge/H-bonding network that engages a fatty acid carboxyl oxygen (labeled “O-I” in Fig. 4)) via the side-chain amide nitrogen of Asn42 (E helix). Ile70 lies at the beginning of the hydrophobic tunnel, making close contact with the first 2 methylenes of the fatty acid. Ala71 points away from the binding pocket, but its short sidechain allows the F helix to pack closely against the H helix, which likely explains its conservation. The residue at position 72 points out into solution, and accordingly is not conserved. Arg74 makes a bifurcated salt-bridge to Asp33 (BE turn), and engages both fatty acid carboxyl oxygens (O-I and O-II) via its positively charged N $\delta$ , forming a direct H-bond to O-II and a water mediated bond to O-I.

A second prominent feature within the hydrophilic chamber of both pXO1-118 and pXO2-61 is a halide ion. In the case of pXO2-61, a very strong electron density peak is observed. An anomalous Difference Fourier identifies the peak as an iodide ion (peak height  $\sim 10 \sigma$ ), since it is the only species present in the protein or crystallization liquor with significant anomalous scattering at the wavelength employed, 0.98 Å (Suppl. Fig. 3). It is the only iodide ion visible within the tunnel or indeed within any buried region of the dimer. We have therefore assigned a strong peak in the analogous position in crystals of pXO1-118 as chloride, since iodide is absent from that crystallization liquor while chloride is abundant, and it has been established in other systems (see below) that Cl<sup>-</sup> and I<sup>-</sup> typically bind competitively to the same sites. Moreover, crystallographic refinement with the appropriate halide ions yielded B values comparable with those of the surrounding protein, consistent with full occupancy of the sites

The environment of the chloride ion is shown in Fig. 4D. Its closest contact (2.6 Å) is made with the second carboxyl oxygen of the fatty acid, “O-II” (suggesting that this oxygen is protonated and forms a H-bond); further contacts are with the side-chain of Arg74, as well as the side-chain amide nitrogen of Asn87. Contacts with the hydrophobic side-chain of Ile39, and one directly-bound water molecule, complete the coordination sphere. The structure of this site is remarkably well preserved in pXO2-61, except that a second water molecule takes the place of the fatty acid, and the larger but more polarizable iodide ion has similar but distinct bond lengths (Table II). It should also be noted that there is an identical chain

of well-defined H-bonded water molecules tracing a narrow hydrophilic channel through the top of the canopy into bulk solvent, suggesting a pathway for halide exchange that does not require release of the fatty acid (Fig. 4B).

The site fits the general characteristics of a buried chloride binding pocket, as described (39). Such sites require sufficient positive potential to neutralize the anion, through Coulombic and/or H-bonding interactions, but the coordination geometry is less stringent than for a typical metal-binding site. Notwithstanding, the site in our sensor domains resembles a well-characterized buried site in the *E. coli* protein, GadB, where chloride plays an allosteric, pH-dependent role (40). Furthermore, the authors in that case showed that Br<sup>-</sup> and I<sup>-</sup> could readily substitute for chloride (see Table II).

The binding pocket has similarities with other fatty acid binding proteins (but created from a different protein architecture). For example, the nuclear receptor HNF4 $\alpha$  (41), which creates a hydrophobic tunnel and buried arginine to engage the fatty acid. The RSCB Protein Data Base also contains an example of a bacterial PAS domain (from the signaling protein, RV1364C, from *Mycobacterium tuberculosis* PDB code: 3K3C) which houses a palmitic acid in a broader pocket. In this case, both an Arg and an Asp (O-O distance = 2.6 Å) coordinate the carboxylic acid, implying that the fatty acid is protonated in this case also. However, no other ions were identified in either pocket.

*pXO1-118 and pXO2-61 bind reversibly to fatty acids in solution.* The fatty acid binding pocket, as well as the dimerization interfaces, are well conserved structurally between pXO1-118 and pXO2-61, and sequence comparisons suggest that the key binding/dimerization determinants of the sensor domains are conserved in BA2291 and indeed throughout the *B. cereus* family of BA2291-related histidine kinases (as well as relatives such as *Geobacillus*) (Fig 5). This suggests that they will all bind the same or a similar ligand and function in a similar way.

The tunnel appears to be optimal for a C14 fatty acid, and there is also a distinct bend in the tunnel around methylene 8, raising the possibility of specificity for a cis-monounsaturated fatty acid (bacteria possess few polyunsaturated fatty acids). We therefore measured the binding of 5 different *Bacillus* fatty acids by ITC. We found that both pXO1-118 and pXO2-61 bound fatty acid reversibly, and measured binding to 2 saturated, unbranched (myristic and palmitic acids), 2

saturated branched (12-methyltetradecanoic and 13-methyltetradecanoic acids), and 1 mono-unsaturated (palmitoleic acid). We found that all bound exothermically with a stoichiometry ~1:1. However, we did not detect significant selectivity among the fatty acids tested (binding affinities were in the range of 10-40  $\mu$ M (Table III and Suppl. Fig. 4).

*A new globin functionality* – Our crystal structures and sequence comparisons illustrate a novel ligand-binding modality for the globin fold, one that does not involve heme or any related porphyrin-like attachment. Structural alignments point to a remarkable equivalence between the proximal histidine of heme proteins and the key arginine residue within the **KIAxER** motif, strongly suggesting an evolutionary relationship. Indeed the heme-based oxygen sensor has the related sequence **KIGHAH**, in which its Ile packs against the hydrophobic moiety of the cofactor (heme), just as the analogous Ile of pXO1-118 does against fatty acid. And the alanine/glycine residue serves a similar purpose, ensuring tight packing between the F and H helices (and providing a potential link between ligand recognition and dimerization).

The fold family containing pXO2-61 and pXO1-118 has been classified by PFAM as “Family: HisK\_N (PF09385)”. The PFAM database currently contains more than 100 sequences, very few of which have been functionally characterized. Many are the sensor domains of *B. cereus* family kinases, which are identical (or nearly so) to BA2291. However, there are also many distinct sequences, which we have sub-classified into 4 groups in light of our structural analysis (see Fig. 5):

*Group I* comprises the “sensor-only” domains from family Bacillus. In addition to pXO1-118 and pXO2-61, a number of *B. cereus* species have acquired a pXO1-like plasmid, and these contain genes that are highly homologous to pXO1-118. The only other sequence is from the recently annotated genome of *B. cellulosilyticus*, an alkaliphilic, salt-tolerant, spore-forming bacterium (42). Its sequence is more divergent, but still demonstrates all of the key structural features associated with the other Group I members. It is not reported whether it is plasmid-encoded.

*Group II* comprises the sensor histidine kinases of the BA2291 family. As noted, these are found in a large number of *B. cereus*/*B. anthracis* strains/isolates, as well as the related *B. thuringiensis*, an insect pathogen. Their sensor domains are nearly invariant (>90-100% identity), so we assume that they are sensing the same or

similar signal. Note that only a small fraction of these species/strains harbor a plasmid-borne sensor-only domain.

*Group III* comprises sensor domains from a more disparate collection of sensor histidine kinases. They are all from the class Bacilli, but from different genera within the family Bacillaceae, including *Geobacillus* and *Brevibacillus*. These typically show around 30-45% identity to the BA2291 class, but again display the key sequence characteristics, suggesting identical or similar functionality.

*Group IV* comprises “sensor-only” domains from bacteria from distinct phyla, for which no cognate kinases have been identified. Most of the organisms bearing these genes occupy distinct or unique environmental niches, and utilize a wide variety of energy sources. For example, the phylum Chlorobi (green sulfur bacteria) is well-represented. These are Gram-negative bacteria that can degrade a wide variety of chloro-aromatic compounds, as well as toxic metals (43,44). The sequences are, not surprisingly, more divergent (~20-25% identity) from those of the Groups I-II, but they still include most of the hallmark residues involved in structural integrity, and, to a lesser extent, cofactor binding and dimerization. They all have Gly or Ala at the 3<sup>rd</sup> position of the motif, consistent with a conserved packing between helices F and H. They typically have a longer BE turn, suggesting a structure more closely related to heme-binding globins. Nothing is known about the function of these domains, but for some members of this Group there is divergence of the KIAxER motif, suggesting related but distinct functions. For example, in an *Anaeromyxobacter* sensor, a histidine replaces the inward-facing Ile, while in an *Acidobacter* sensor, the Arg at position 6 is replaced by cysteine. Both changes introduce sidechains with the potential to generate novel functionalities within the central cavity.

## DISCUSSION

We have identified 2 ligands that bind to pXO1-118 and pXO2-61 sensor domains *in vitro* - fatty acid and halide (chloride) – and our data strongly suggest that the chromosomally encoded kinase, BA2291, behaves similarly, and will therefore sense the same environmental cue *in vivo*, thus supporting the proposal that the plasmid-encoded sensor domains inhibit untimely sporulation by titrating the signal (22). Our studies point to roles for fatty acid (or similar molecule), chloride ion, and possibly pH (see below), as signaling cues,

which should inform the next set of experiments to define their activity *in vivo*.

Fatty acid synthesis is upregulated in preparation for sporulation, both in quantity and type, with a shift toward monounsaturated species. Thus, the sensor domains might recognize a fatty-acid that is synthesized in the build-up to sporulation. Our short survey of saturated and unsaturated fatty acids did not offer any clues in this regard, as they all bound with similar affinity. Nevertheless, the binding pocket we have described does have a distinct shape that could be optimized for a specific unsaturated species. Specificity of this kind has been observed for an unrelated sensor kinase, DesK, from *B. subtilis* (45), which recognizes C16 fatty acids with a double bond at the  $\Delta 5$  position.

Chloride ion concentrations differ by 10-20 fold between the different host tissues that *B. anthracis* must encounter during pathogenesis - from the intracellular milieu of the macrophage (~5 mM) to the extracellular fluids (lymphatic and plasma >100 mM) (46) - which could reasonably offer a trigger for sporulation (that needs to be suppressed). We have demonstrated the existence of a specific chloride ion binding site that is intimately involved in the binding of fatty acid, raising the possibility that chloride (or some other small anion) is the “ligand” and fatty acid the “cofactor”, by analogy with oxygen binding to heme in the hemoglobins. We note that the ligand binding geometry is consistent with the fatty acid being protonated, enabling one carboxyl oxygen (O-II) to form a favorable hydrogen bond with the chloride ion. Thus, fatty acid binding might be

both chloride- and pH-dependent within the physiological range (*cf.* the Bohr/chloride effects in hemoglobin (47)) and the combinatorial effects of chloride and pH on the activity of *E. coli* decarboxylase, GadB (40)). Furthermore, significant changes in environmental pH are associated with different steps in *B. anthracis* pathogenesis, any of which could, in principle, offer a trigger for sporulation. However, further work is clearly required to test this hypothesis.

The broader significance of our work lies in the definition of a new family of globin-based sensor domains that utilize an alternative (non-heme) cofactor, but which nevertheless appear to be closely linked evolutionarily to the heme-bearing globins. The majority of sensor domains within this family (that have been sequenced so far) come from genomes within the family Bacillaceae, and are contained within a histidine kinase architecture that is very similar structurally and most likely functionally to the *B. anthracis* BA2291 sporulation kinase. By contrast, recent genome sequencing efforts have uncovered a sub-family of “sensor-only” domains that are found in a range of bacteria from unrelated phyla. In general, rather little is known about these bacteria, other than a shared propensity for utilizing unusual energy sources such as the hydrolysis of chloroaromatics. Nothing is currently known about the function of their putative sensor domains, but it will be most interesting to see if they play a role in regulating these unusual metabolic functions.

## REFERENCES

1. Okinaka, R.T., Cloud, K., Hampton, O., Hoffmaster, A.R., Hill, K.K., Keim, P., Koehler, T.M., Lamke, G., Kumano, S., Mahillon, J., Manter, D., Martinez, Y., Ricke, D., Svensson, R., Jackson, P.J. (1999) *Journal of Bacteriology* **181**, 6509-6515
2. Ezzell, J. W., and Welkos, S. L. (1999) *J Appl Microbiol* **87**, 250
3. Candela, T., Mock, M., and Fouet, A. (2005) *J Bacteriol* **187**, 7765-7772
4. Mock, M., and Fouet, A. (2001) *Annu Rev Microbiol* **55**, 647-671
5. Kang, T. J., Fenton, M. J., Weiner, M. A., Hibbs, S., Basu, S., Baillie, L., and Cross, A. S. (2005) *Infect Immun* **73**, 7495-7501
6. Drysdale, M., Heninger, S., Hutt, J., Chen, Y., Lyons, C. R., and Koehler, T. M. (2005) *EMBO J* **24**, 221-227
7. Hu, H., Sa, Q., Koehler, T. M., Aronson, A. I., and Zhou, D. (2006) *Cell Microbiol* **8**, 1634-1642
8. Perego, M., and Hoch, J. A. (2008) *Trends Microbiol* **16**, 215-221
9. Dai, Z., Sirard, J. C., Mock, M., and Koehler, T. M. (1995) *Mol Microbiol* **16**, 1171-1181
10. Uchida, I., Hornung, J.M., Thorne, C.B., Klimpel, K.R. and Leppla, S.H. (1993) *J Bacteriol.* **175**, 5329-5338
11. Drysdale, M., Bourgogne, A., Hilsenbeck, S.G., Koehler, T.M. (2004) *Journal of Bacteriology* **186**, 307-315



12. Bourgogne, A., Drysdale, M., Hilsenbeck, S.G., Peterson, S.N., Koehler, T.M. (2003) *Infect. Immun.* **71**, 2736-2743
13. Mignot, T., Fouet, A. and Mock, M. (2003) *Mol. Microbiol.* **47**, 917-927
14. Hadjifrangiskou, M., Chen, Y., and Koehler, T. M. (2007) *J Bacteriol* **189**, 1874-1883
15. Saile, E., and Koehler, T. M. (2002) *J Bacteriol* **184**, 370-380
16. Stephenson, K., and Hoch, J. A. (2002) *Mol Microbiol* **46**, 297-304
17. Brunsing, R. L., La Clair, C., Tang, S., Chiang, C., Hancock, L. E., Perego, M., and Hoch, J. A. (2005) *J Bacteriol* **187**, 6972-6981
18. Bongiorno, C., Stoessel, R., Shoemaker, D., and Perego, M. (2006) *J Bacteriol* **188**, 487-498
19. Bongiorno, C., Stoessel, R., and Perego, M. (2007) *J Bacteriol* **189**, 2637-2645
20. Szurmant, H., and Hoch, J. A. (2010) *Curr Opin Microbiol*
21. Scaramozzino, F., White, A., Perego, M., and Hoch, J. A. (2009) *J Bacteriol* **191**, 687-692
22. White, A. K., Hoch, J. A., Grynberg, M., Godzik, A., and Perego, M. (2006) *J Bacteriol* **188**, 6354-6360
23. Harrison, C. J., Bohm, A.A., Nelson, H.C.M. (1994) *Science* **263**, 224-227
24. Otwinowski, Z., Minor, W. (1997) *Methods Enzymol.* **276**, 307-326
25. Terwillinger, T. C., Berendzen, J. (1999) *Acta Crystallogr D Biol Crystallogr* **55**, 849-861
26. Terwillinger, T. C. (2001) *Acta Crystallogr D Biol Crystallogr* **57**, 1755-1762
27. Kleywegt, G. J., Zou, J.Y., Kjeldgaard, M. and Jones, T.A. (2001) Around O. in *International Tables for Crystallography, Crystallography of Biological Macromolecules* (Rossmann, M. G. A., E., ed.), Dordrecht: Kluwer Academic Publishers, The Netherlands. pp 353-356 and 366-367
28. Murshudov, G. N., Vagin, A.A., Dodson, E.J. (1997) *Acta Crystallogr D Biol Crystallogr* **53**, 240-255
29. Brünger, A. T., Warren, G.L. (1998) *Acta Crystallogr D Biol Crystallogr* **54**, 905-921
30. Laskowski, R. A., MacArthur, M. W., Moss, D. S., Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26**, 283-291
31. Holm, L., Sander, C. (1993) *J. Mol. Biol.* **233**, 123-138
32. Murray, J. W., Delumeau, O., and Lewis, R. J. (2005) *Proc Natl Acad Sci U S A* **102**, 17320-17325
33. de Sanctis, D., Dewilde, S., Pesce, A., Moens, L., Ascenzi, P., Hankeln, T., Burmester, T., and Bolognesi, M. (2004) *J Mol Biol* **336**, 917-927
34. Tarricone, C., Galizzi, A., Coda, A., Ascenzi, P., Bolognesi, M. (1997) *Structure* **5**, 497
35. Contreras-Martel, C., Matamala, A., Bruna, C., Poo-Caamano, G., Almonacid, D., Figueroa, M., Martinez-Oyanedel, J., and Bunster, M. (2007) *Biophys Chem* **125**, 388-396
36. Zhang, W., Phillips, G. N. Jr. (2003) *Structure* **11**, 1097-1110
37. Kaneda, T. (1967) *Journal of Bacteriology* **93**, 894-903
38. Westh, P., Kato, H., Nishikawa, K., and Koga, Y. (2006) *J Phys Chem A.* **110**, 2072-2078
39. Song, Y., and Gunner, M. R. (2009) *J Mol Biol* **387**, 840-856
40. Gut, H., Pennacchietti, E., John, R. A., Bossa, F., Capitani, G., De Biase, D., and Grutter, M. G. (2006) *EMBO J* **25**, 2643-2651
41. Dhe-Paganon, S., Duda, K., Iwamoto, M., Chi, Y. I., and Shoelson, S. E. (2002) *J Biol Chem* **277**, 37973-37976
42. Nogi, Y., Takami, H., and Horikoshi, K. (2005) *Int J Syst Evol Microbiol* **55**, 2309-2315
43. Sakurai, H., Ogawa, T., Shiga, M., and Inoue, K. (2010) *Photosynth Res* **104**, 163-176
44. Perez-Pantoja, D., De la Iglesia, R., Pieper, D. H., and Gonzalez, B. (2008) *FEMS Microbiol Rev* **32**, 736-794
45. Aguilar, P. S., Hernandez-Arriaga, A.M., Cybulski, L.E., Erazo, A.C., de Mendoza, D. (2001) *The EMBO Journal* **20**, 1681-1691
46. Lodish, H., Baltimore, D., Berk, A., Zipursky, L. S., Matsudaira, P., and Darnell, J. (1995) *Molecular Cell Biology*, 3rd ed., W. H. Freeman and Company, New York
47. Giardina, B., Mosca, D., and De Rosa, M. C. (2004) *Acta Physiol Scand* **182**, 229-244

## FOOTNOTES

<sup>4</sup>Present Address: Department of Molecular Biology, University of Salzburg, 5020 Salzburg, Austria.

<sup>5</sup>The abbreviations used are: ITC, Isothermal Titration Calorimetry; AUC, analytical ultracentrifugation; GC-MS, Gas Chromatography-Mass Spectrometry.

## ACKNOWLEDGEMENTS

This work was supported by NIH (AI055789 and AI055860). We thank Annie Heroux for collecting diffraction data at the National Synchrotron Light Source, Brookhaven National Laboratory, which is supported by the U.S. Department of Energy (DOE), Division of Materials Sciences and Division of Chemical Sciences, under Contract No. DE-AC02-98CH10886. Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. We thank the staff of the SSRL Structural Molecular Biology Program, which is supported by the DOE, Office of Biological and Environmental Research, and by the NIH, NCRR, Biomedical Technology Program, and the NIGMS. G.S. received an Erwin Schrödinger fellowship from the Austrian Science Fund. M.G. was supported in part by grant 2 PO4B 026 28 from the Polish State Committee for Scientific Research.

## FIGURE LEGENDS

**Fig. 1.** Structure of the *B. anthracis* pXO1-118 sensor domain. *A.* Ribbon representation of the dimer (side view), with termini and helices labeled (according to standard globin nomenclature). Helices are colored in spectral order (blue-> orange) for each monomer. Labels for second monomer are primed. The KIAxER motif within helix F is colored in red. Fatty acid is shown as cyan (methylene carbons) and red (carboxyl oxygens) spheres. *B.* Same as in *A*, but orthogonal view looking down the 2-fold axis of the dimer. *C.* Same view as in *A*, highlighting the dimer interface, with water molecules found only in a central region. *D.* Sequence alignments of pXO1-118, pXO2-61, and the sensor domain of the *B. anthracis* sporulation kinase, BA2291. Secondary structure elements for pXO1-118 are indicated, as is the KIAxER motif.

**Fig. 2.** Stereo C $\alpha$  overlay of pXO1-118 (red) and pXO2-61 (gray) dimers, with N and C-termini labeled. Occasional residue numbering is given for guidance. The fatty acid in crystals of pXO1-118 is shown in cyan (methylene carbons) and red (carboxyl oxygens).

**Fig. 3.** pXO1-118 has a globin fold but does not bind heme. *A.* Overlay of pXO1-118 monomer (red) with the heme-based sensor from *Geobacter sulfurreducens* (blue). Heme and heme-linked proximal histidine are shown for the latter. Helices A, B, F and the FG turn, G (not visible) and H overlap closely, while for pXO1-118, helix E tilts and bends toward helix F, occluding the heme pocket (fatty acid has been omitted for clarity). *B.* The same overlay as in *A* (rotated ~90° about a vertical axis), showing the FG region in greater detail. Note how well the main-chain aligns through the FG turn. The C $\alpha$  of residue F8 is labeled. In the heme-based sensor, this residue is the “proximal histidine”, which engages the heme-linked iron atom (gray sphere). In pXO1-118, this residue is Arg74, which engages the head-group of the fatty acid (carboxyl oxygens are in red), which nearly coincides with the position of the heme-linked iron. The chloride ion (green) also lies nearby. *C.* Cut-away view of the cavity at the center of pXO1-118 (view is similar to *B*), showing that its shape is incompatible with heme binding. Note that the hydrophilic propionate sidechains (on the right side of the heme) are confined within the protein (in this hypothetical model), owing to the close packing of helices E and F. *D.* Structure-based sequence comparisons of the F and G helices of the sensor domains and RsbR, which does not bind cofactor, illustrating the truncated F helix that lacks an “F8” residue (the standard globin numbering scheme is

given for the F helix). The conserved KIAxER motif is boxed. The heme-based sensor is also aligned to illustrate its related motif.

**Fig. 4.** Sensor domain interactions with fatty acid and chloride. *A.* Slice through the 2Fo-Fc electron density map of pXO1-118 showing fatty acid (carbon in blue, carboxy oxygens in red, labeled O-I and O-II (see text)), chloride ion (green), and water molecules (red balls) in the context of the hydrophilic chamber. *B.* Model of the same region as in *A*, showing the salt-bridge and H-bonding network comprising the “canopy” that encloses the hydrophilic chamber. Note the network of water molecules (circled and labeled I-III) that trace a potential exit route for the chloride ion. Helices and turns are labeled (and circled) for guidance. The orientation of the amide sidechain of Asn87, which plays a key role in chloride and fatty acid binding, is unambiguous, as it is defined by 2 tandem H-bonds to the mainchain of Leu28. *C.* Conserved interactions between the methylene carbons of the fatty acid and the hydrophobic/aromatic sidechains lining the walls of the tunnel. Tetradecanoic acid has been modeled. The sidechains of conserved residues are shown in green. The Ca atoms of 2 conserved glycines (Gly43 and Gly91) that delimit the cavity width in that region, are also labeled. Trp23 at the back of the pocket is invariant in all *Bacillus* sensor domains. Phe19 is the “gatekeeper”, shown in the “open” configuration. *D.* Stereo view of the chloride ion coordination site, showing H-bonding, Coulombic and van der Waals interactions with protein and fatty acid (the O-II carboxyl oxygen is labeled). Bond distances are provided in Table III, and compared with an *E. coli* GadB regulatory site.

**Fig. 5.** Conservation of structure and function in the sensor domain family. *A.* Structure-based sequence alignment and conservation of functional surfaces: fatty acid binding pocket (hydrophobic tunnel), highlighted in green; canopy creating fatty acid/chloride binding site (blue); dimer interface (yellow); and other conserved residues important for structural integrity (cyan). Sequences are grouped into 4 classes (I-IV) as defined in the main text. Full names of the organisms and genes are as follows. **GROUP I:** *B. anthracis* pXO1-118 (NP\_052814), *B. anthracis* pXO2-61 (NP\_653040), *B. cereus* G2941 (ZP\_00236329), *B. cellulosilyticus* DSM 2522 hypothetical protein BcellDRAFT\_3961 (ZP\_06365458). **GROUP II:** *B. anthracis* BA2291 (NP\_844676.1), *B. cereus* sensor histidine kinase (YP\_083662.1), *B. thuringiensis* sensor histidine kinase (YP\_036402.1). **GROUP III:** *B. coahuilensis* m4-4 sensor histidine kinase (ZP\_03225478), *Bacillus* sp. SG-1 histidine protein kinase (ZP\_01860026), *Brevibacillus brevis* NBRC 100599 probable two-component sensor histidine kinase (YP\_002774563), *Geobacillus thermoglucosidasius* C56-YS93 histidine kinase (ZP\_06810042), *Geobacillus* sp. C56-T31 histidine kinase (YP\_003671552), *Bacillus* sp. NRRL B-14911 sensor histidine kinase (ZP\_01170643), *Bacillus* sp. SG-1 (ZP\_01860665); **GROUP IV:** *Anaeromyxobacter dehalogens* 2CP-C conserved hypothetical protein (ACL66915.1), *Chlorobium luteolum* DSM 273 hypothetical protein Plut\_0146 (ABB23036.1), *Chloroherpeton thalassium* ATCC 35110 conserved hypothetical protein (ACF14428.1), *Syntrophus aciditrophicus* SB hypothetical cytosolic protein (ABC79005.1), Candidatus *Solibacter usitatus* Ellin6076 hypothetical protein Acid\_3380 (ABJ84353.1). *B.* Stereo image of a pXO1-118 monomer, together with the dimer interface of the second monomer. Main-chain (N-C $\alpha$ -C) is shown as ribbon. The C $\alpha$  positions of residues highlighted in *A* are shown as spheres, using the same color scheme. Fatty acid is in cyan and red, chloride ion in magenta.

**Table I** X-ray data collection, phasing and refinement.

	Phasing	Model refinement	
	$\lambda_{\text{peak}}$	<b>pXO1-118</b>	<b>pXO2-61</b>
Space Group		P3 <sub>2</sub> 21	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Cell Dimensions (Å)		a=89.9 c=35.3	a=44.1 b=62.6 c=124.7
Wavelength (Å)	0.9781	0.97923	0.97923
Resolution range (Å)	50-2.5	80-1.76	60-1.49
Observations	70890	173637	408633
Unique reflections	5888	16461	56717
Completeness <sup>1</sup> (%)	99.8 (100)	99.5 (95.0)	98.9 (94.5)
R <sub>SYM</sub> <sup>1,2</sup> (%)	6.9 (25)	5.7 (46)	6.8 (32)
R <sub>CRYST</sub> <sup>3</sup> /R <sub>FREE</sub> <sup>4</sup> (%)		.185/.241	.177/.209
Protein atoms		1408	2628
Water molecules		95	364
Other ions		1	34
Ligand		1	
<b>RMS deviation:</b>			
Bonds (Å)		0.027	0.012
Angles (°)		1.70	1.37
<b>Average B-factor (Å<sup>2</sup>):</b>			
Main chain		25.0	18.9
Side chain		32.7	24.0
Water		34.6	35.9
Ligands/Ions		37.4	39.7

<sup>1</sup>Numbers in parentheses are for highest resolution shell. <sup>2</sup>R<sub>SYM</sub>= $\sum |I_h - \langle I_h \rangle| / \sum I_h$ , where  $\langle I_h \rangle$  is the average intensity over symmetry equivalent reflection. <sup>3</sup>R-factor =  $\sum |F_{\text{obs}} - F_{\text{calc}}| / \sum F_{\text{obs}}$ , where summation is over the data used for refinement. <sup>4</sup>R<sub>FREE</sub> was calculated using 5% of data excluded from refinement.

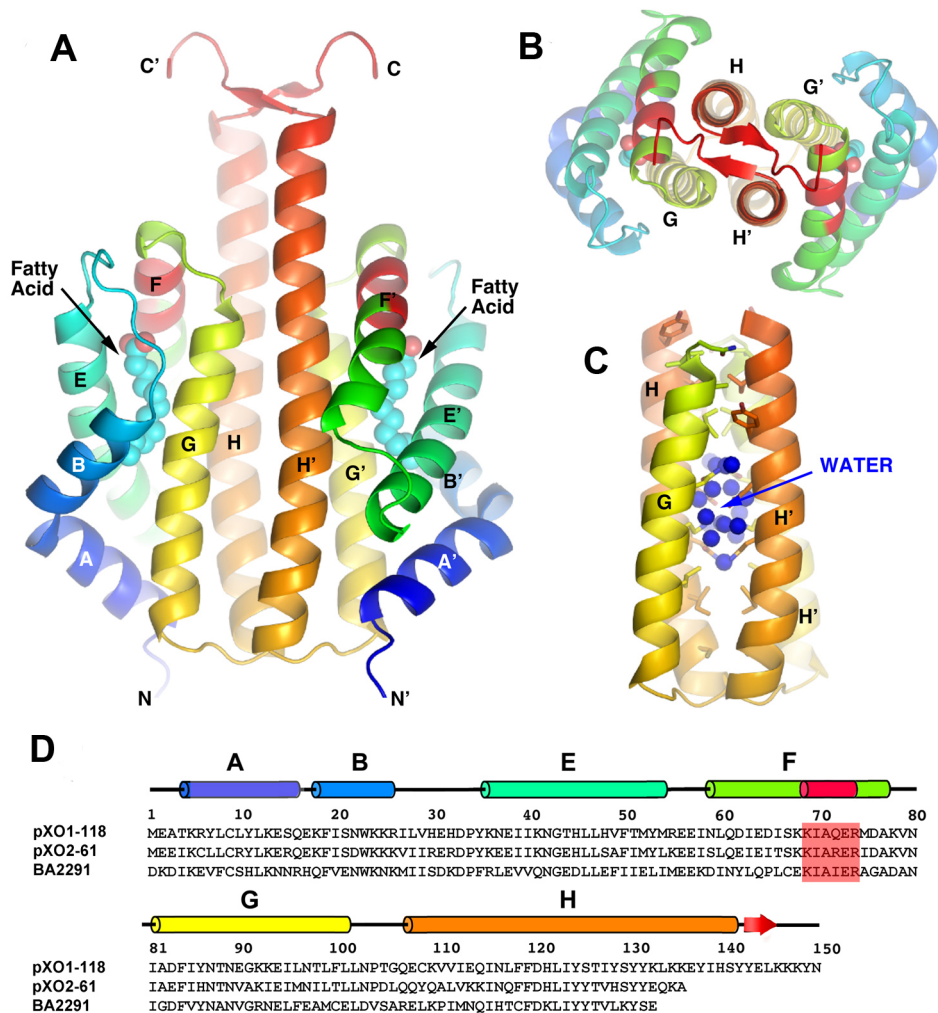
**Table II.** Environments of the halide ions and comparison with an allosteric site in *E. coli* GadB.

<b>118/61</b>	<b>Cl<sup>-</sup> / Å</b>	<b>I<sup>-</sup> / Å</b>	<b><i>E.coli</i> GadB</b>	<b>I<sup>-</sup> / Å</b>
Arg74 C $\gamma$ /N $\delta$	4.6/4.2	4.4/4.1	Arg17 NH/C $\epsilon$	3.5/3.7
84NH	4.1	4.6	17NH	4.0
Asn87 N $\delta$ 2	3.2	3.7	18NH	4.3
Water	3.1	3.5, 3.6	Water	3.7
Ile39 s/c	4.3	4.1	Trp s/c	4.1
Fatty acid –CO <sub>2</sub> H	2.8	3.4	Ser-OH	3.2

Interatomic bond distances are given in Å; see Fig. 4D. s/c=sidechain.

**Table III.** Binding affinities (Kd) of pXO2-61 and pXO1-118 for selected fatty acids.

Fatty Acid	pXO2-61 Kd / $\mu$ M	pXO1-118 Kd / $\mu$ M
Myristic acid	40 $\pm$ 12	25 $\pm$ 9
Palmitic acid	41 $\pm$ 17	24 $\pm$ 7
12-Methyltetradecanoic acid	41 $\pm$ 19	14 $\pm$ 8
13-Methyltetradecanoic acid	46 $\pm$ 15	13 $\pm$ 6
Palmitoleic acid	20 $\pm$ 10	30 $\pm$ 8



**Fig. 1**



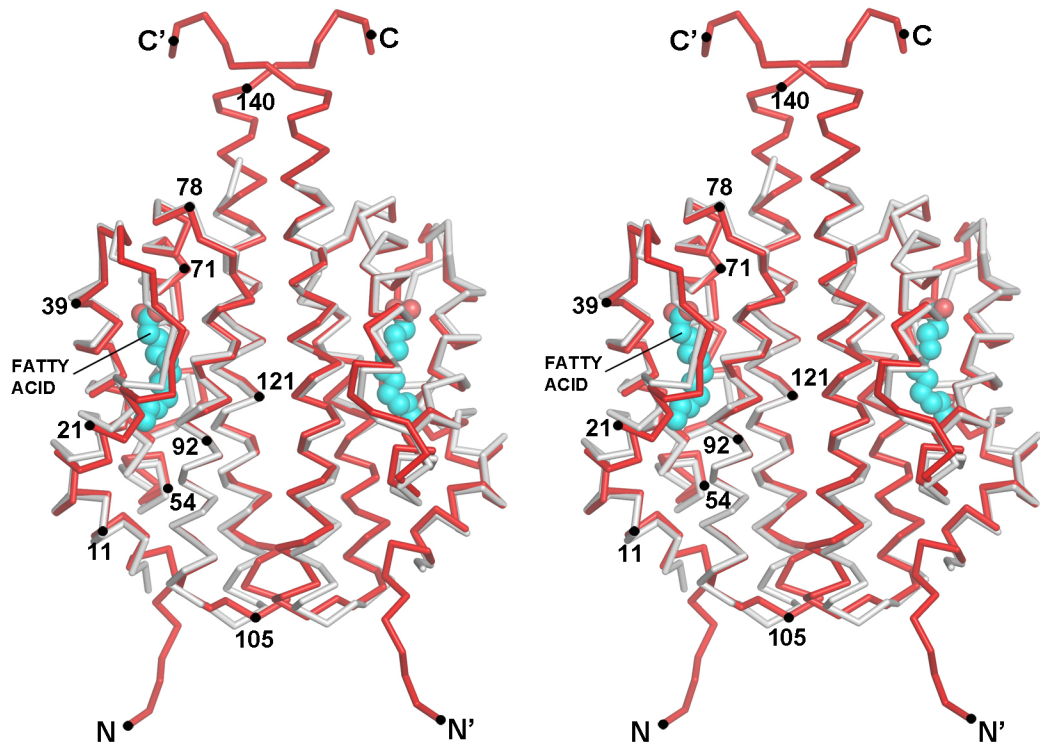
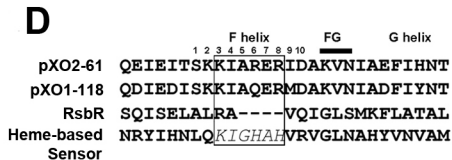
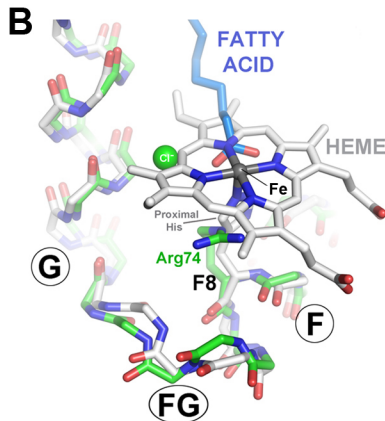
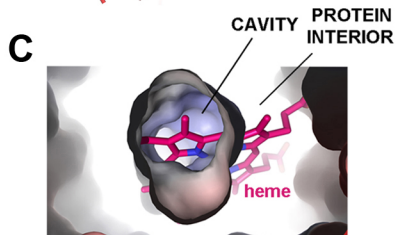
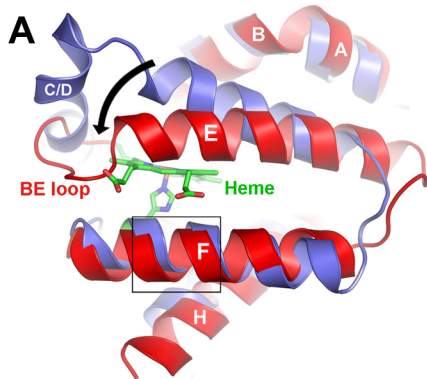
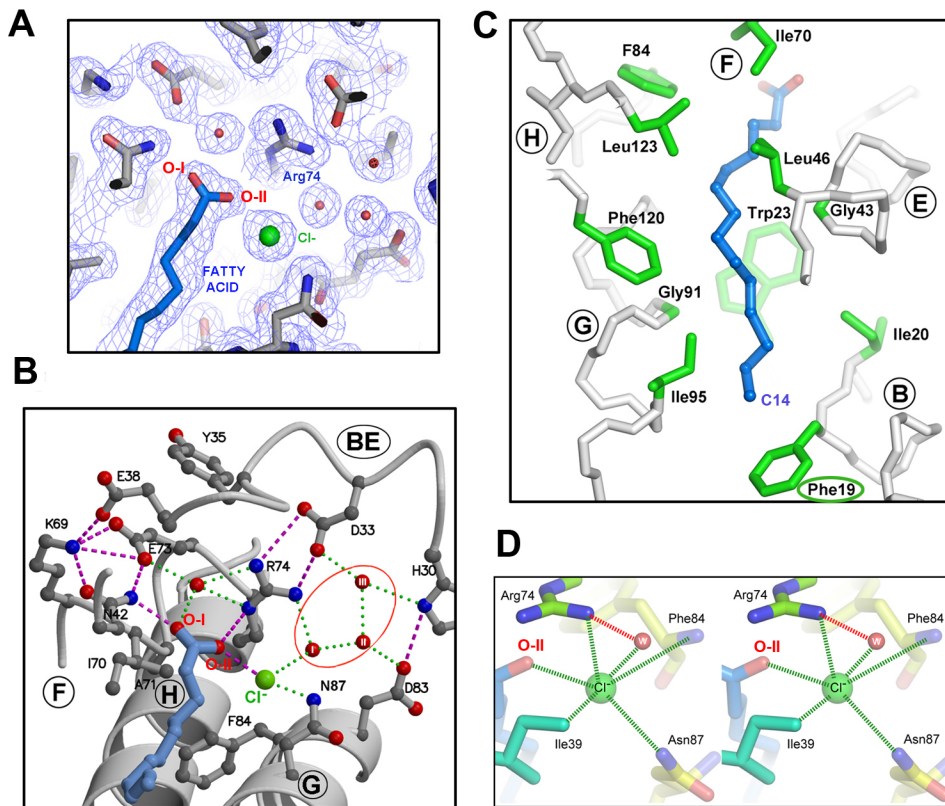


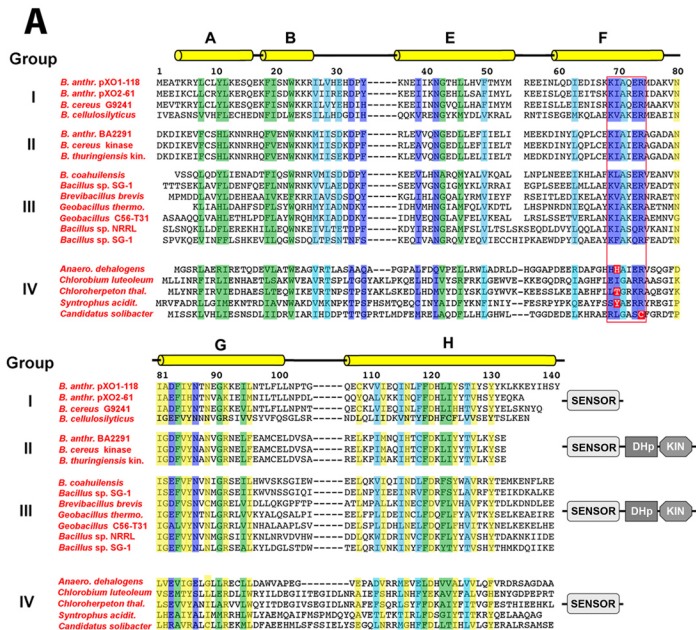
Fig 2



**Fig. 3**



**FIG 4**



**Fig. 5**